ChIP-Seq

Q1. What is meant by ChIP-Seq?

A1. From Illumina [1]: "*Chromatin immunoprecipitation (ChIP) is a powerful method to selectively enrich for DNA sequences bound by a particular protein in living cells. ChIP-Seq on Illumina sequencing systems supports virtually unconstrained selection of any ChIP-able protein and/or modification to be studied. These include transcription factors, polymerases and transcriptional machinery, structural proteins, protein modifications, and DNA modifications. … The ChIP process enriches specific crosslinked DNA protein complexes using an antibody against a protein of interest. Unique oligonucleotide adapters are then added to the small stretches of DNA that are bound to the protein of interest to enable massively parallel sequencing.*" Some applications of this technology include [2]:

- Discovery of novel transcripts and transcription factor binding sites
- Identification of genes regulated by known transcription factors and co-regulators
- Analysis of epigenetic events
- Direct comparison of regulatory events in different cell states (i.e. normal vs. disease)
- Investigation of drug effects and other stimuli on regulatory pathways

Q2. What material should I send to be analyzed by ChIP-Seq?

A2. Generally, we like to start with 200 ng of ChIP-enriched DNA. If this much material is unattainable we will accept less, but this significantly reduces the chances of obtaining a representative library. The sample should be QC'd by the investigator by looking for the relative enrichment of a relevant gene. The best control material is an unprocessed aliquot of the input DNA that went into the ChIP enrichment step. A light sequencing of this sample can reveal potential false-positives.

Q3. What data are returned by NISC?

A3. Typically, NISC returns to the investigator a file containing basecalls and quality scores (BAM files). The investigator is expected to provide data analyses; this is not offered by NISC. Data files can be quite large. Each lane can produce 20-30 million single-end reads.

Q4. How does ChIP-Seq compare to ChIP-Chip?

A4. NISC has not performed this comparison, but Illumina provides the following list of features [3]:

- High-Quality Data: Positional resolution of mapped binding sites ± 5 bp
- Wide Dynamic Range: Robust quantification for determining binding specificities of varying strengths
- High Signal-to-Noise Ratio: Lower background than ChIP-chip, no cross hybridization
- Genome-Wide Analysis: Identifies any binding sites, not limited to array features or candidate sequences

Q5. How long do the reads need to be for ChIP-Seq analysis?

A5. Typically, fragment read lengths are 35 bases. This length should be sufficient for mapping of most reads to the reference genome. Some investigators are exploring the utility of longer reads and paired-end reads for advanced analyses.

Q6. How many lanes of a flow cell are used for a mammalian ChIP-Seq analysis?

A6. Each lane of a flow cell yields 20-30 million single-end reads. A single lane will be sufficient in ChIP-Seq experiments of low complexity, where a small fraction of the genome is enriched. An additional lane may be required in experiments which are enriched for a large fraction of the genome.

References:

1. Illumina, Inc. (2011): ChIP-Seq DNA Sample Prep Kit
http://www.illumina.com/products/chip-seq_dna_sample_prep_kit.ilmn
2. Beckman Coulter Genomics (2011):
http://www.beckmangenomics.com/genomic_services/dna_sequencing/chip_seq.html
3. Illumina, Inc. (2011): ChIP-Seq Assay
http://www.illumina.com/technology/chip_seq_assay.ilmn